DOCUMENT RESUME

ED 476 919 TM 034 970

AUTHOR	Newman, Isadore; Hall, Rosalie J.; Fraas, John
TITLE	Development of a Regression Model for Estimating the Effects of Assumption Violations on Type I Error Rates in the Student's T-Test: Implications for Practitioners.
PUB DATE	2003-04-00
NOTE	25p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 21-25, 2003).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Estimation (Mathematics); Monte Carlo Methods; Multiple Regression Analysis; *Regression (Statistics)
IDENTIFIERS	*T Test; Type I Errors; *Violation of Assumptions

ABSTRACT

Multiple linear regression is used to model the effects of violating statistical assumptions on the likelihood of making a Type I error. This procedure is illustrated for the student's t-test (for independent groups) using data from previous Monte Carlo studies in which the actual alpha levels associated with violations of the normality assumption, homogeneity of variance, or unbalanced designs were determined. The observed Type I error rates were recorded, along with information coding the type and extent of statistical assumption violation. The resulting linear models had R squared values of 0.88 to 0.91 and adjusted R squared values of 0.87 to 0.90. The results of the suggested methodological approach: (1) reveal the feasibility of developing multiple linear regression models to predict actual Type I error rates based on various assumption violation conditions for the independent groups-t-test; (2) suggest that alpha inflation is rarely larger than a factor of 2; and (3) provide a template for the development of assumption violation models for other types of statistical tests. (Contains 3 tables and 27 references.) (Author/SLD)

Running head: Regression Model for Assumption Violation

Development of a Regression Model for Estimating the Effects of Assumption Violations on Type I Error Rates in the Student's t Test: Implications for Practitioners

Isadore Newman

Educational Foundations and Leadership

University of Akron

Rosalie J. Hall

Department of Psychology

University of Akron

John Fraas

Business Administration

Ashland University

Paper presented at the 84th Annual Meeting of the American Educational Research Association, Chicago, IL, April 21-25, 2003

Abstract

Multiple linear regression is used to model the effects of violating statistical assumptions on the likelihood of making a Type I error. This procedure is illustrated for the student's t-test (for independent groups), using data from previous Monte Carlo studies in which the actual alpha levels associated with violations of the normality assumption, homogeneity of variance, and/or unbalanced designs were determined. The observed Type I error rates were recorded, along with information coding the type and extent of statistical assumption violation. The resulting linear models had an R^2 values of .88 to .91 and adjusted R^2 values of .87 to .90. The results of the suggested methodological approach: (a) reveal the feasibility of developing multiple linear regression models to predict actual Type I error rates based on various assumption violation conditions for the independent groups t-test; (b) suggest that alpha inflation is rarely larger than a factor of 2; and (c) provide a template for the development of assumption violation models for other types of statistical tests.

Development of a Regression Model for Estimating the Effects of Assumption Violations on Type I Error Rates in the Student's t-Test: Implications for Practitioners

In spite of the development of a large number of non-parametric alternatives, much educational and psychological research currently relies heavily on a small number of parametric statistical procedures such as t-tests, analysis of variance (ANOVA) and correlation and regression analyses. Typically, when researchers learn to apply these techniques they also learn a set of underlying statistical assumptions that must be met before use of the statistical test is justified. These include assumptions such as the normality of the distribution of the dependent variable, the assumed homogeneity of error variances, and homoskedasticity, as well as other types of assumptions such as the independence of observations or linearity of relationships. Violating such assumptions is known to result in biased parameter estimates and incorrect probability values, although in many real-life situations the assumptions may be closely approximated (Hays, 1994; Maxwell & Delaney, 2000) and there is a historical tradition in many disciplines of relying on normal-theory-based statistics (Huberty, 1987).

However, our own accumulated experiences with academics, other professionals who use statistics, and graduate students, suggest to us that many researchers either: (a) fail to test distributional assumptions altogether, so do not even realize the possibility that their chosen analysis will be in error; (b) realize that their data violate assumptions but use default analytic options which assume no violations anyway, believing that the consequences are negligible; or (c) acknowledge the need to correct for violated assumptions, but choose not to do so because of perceived practical roadblocks (e.g., lack of access to the proper statistical correction or beliefs that the hassles of dealing with reviewers and readers unfamiliar with the technique outweigh its advantages).

Given this state of affairs, one wonders how much confidence can be placed in the body of empirically-based knowledge that has accumulated on many research topics. If at least some degree of violation is present in many of our studies, how confident can we be in the conclusions based on our empirical results? In some types of applications, a meta-analysis of the relevant content studies, with its focus on estimation of effect sizes, may help to address this question. However, in some other situations (especially applied situations where the issue being addressed is very context-specific), there may be insufficient studies available for meta-analysis, or a need to make a decision based on a single isolated study. A key concern in such situations might simply be whether or not the Type I error rate has been substantially inflated by the violation of assumptions, thus resulting in an unacceptable likelihood of concluding that an effect exists, when in fact it does not. In other words, we might find it very helpful to have information that allows us to decide whether the alpha level in such a study should be taken at face value.

The current paper represents our initial attempts to demonstrate a technique that has the potential to estimate (for given statistical analysis in the context of a single study) the likely effects on the Type I error rate of a given constellation of assumption violations. More specifically, linear regression is used to model effects of systematically varying the type and degree of assumption violations on Type I error rates, using simulation studies with known population and sampling conditions as data sources for the regression models. The linear regression model incorporates both the main effects of violating such assumptions as normality and homogeneity of variance, as well as the interactive effects of multiple violations. Once such a model is carefully developed, it can then be used to estimate the likely extent of bias in Type I error rate for a specific given set of assumption violations.

This approach is applied to the case of the robustness of the independent groups student t-test (see Harwell, Rubinstein, Hayes, & Olds, 1992 for a similar application to one- and two-factor fixed effects ANOVA). One factor considered in our investigation is whether the homogeneity of variance assumption is violated, along with additional information about whether or not the experimental design is balanced. Previous work has established that violations of homogeneity of variance have less serious consequences for alpha when the sample sizes of the two groups are equal, but may be quite pronounced when heterogeneity accompanies unbalanced sample sizes (e.g., Boneau, 1960; Glass, Peckham, & Sanders, 1972; Norton, 1952; Scheffe, 1959; Young & Veldman, 1963). The specific implication for alpha depends upon the pattern of sample size and variance inequality when the smaller sample *n* is associated with the larger population variance (i.e., an inverse relationship), alpha is inflated; when the smaller sample *n* is associated with the smaller population variance (i.e., a direct relationship), the actual alpha may be substantially lower than the nominal significance level (Glass et al., 1972).

In addition, the normality assumption is considered. Skewness and kurtosis will affect the actual alpha level (Pearson, 1931; Scheffe, 1959), and the effects of skewness may be more pronounced when the two samples are drawn from populations which are skewed in opposite directions (Tiku, 1964). However, violations of the normality assumption tend to have a smaller effect than heterogeneity, especially when sample sizes are moderately large to large. Also considered is the effect of having extreme outliers present in the distribution.

In sum, the data used as input to the regression model are from a meta-analytic review of the results of several previously-published Monte Carlo studies—many of which were undertaken to demonstrate the biasing effects on the t-test of failing to use the appropriate statistical correction and to contrast the uncorrected t-test results with some preferred alternative analysis! However, our emphasis is different than that of the original investigators in these studies. In addition to providing a demonstration of a more general technique, we hope to use the data to establish some rough boundary conditions that will help practitioners both in evaluating their confidence in previous research using t-tests, and guide them in their choice of current data analytic technique. For example, our results might allow practitioners to make judgments about when the likely effects of violating assumptions justify costs associated with using specialized corrections (e.g., extra time, effort, and loss of easy interpretability and communication) and when the likely effects of violating assumptions will only minimally affect the outcomes and interpretations.

Methodology

Definition of problem. The approach taken was very similar to that recommended by Cooper (1982) and Harwell (1992). We first defined the specific research question that we wished to address. The student's t-test for independent groups was chosen because the use of this test depends upon three statistical assumptions common to many parametric statistical techniques, yet there are a limited number of variations on the basic experimental design for the t-test, thus simplifying the task of providing an illustration. The three relevant statistical assumptions are: (a) the normality of the underlying population distributions from which each group was sampled; (b) equal variances in the two population distributions; and (c) independence of observations. In this illustration, we chose to focus on the first two assumptions, but note that it has been demonstrated that violations of the independence assumption clearly may bias the results of an ANOVA, and by extension, the results of an independent groups t-test (e.g., Kenny & Judd, 1986).

Identification of studies. Accessible studies which reported Monte Carlo simulation data relevant to the research question were identified. These were studies in which: (a) the focal analyses involved the comparison of two independent group means, and the statistical tests used for the analyses included student's t-tests or a two-groups fixed-effects ANOVA; (b) at least some of the samples were non-normal, had heterogeneous variances, and/or unequal sample sizes; and (c) Type I error rates were reported. In all cases, the data sets appeared to have been generated using random sampling, and the independence assumption was not violated. In most of the studies, the data sets were generated with a population mean difference (delta) of zero.

The search yielded a total of nine data sources which reported the observed and nominal alphas, sample size, form of the population distributions, and variance homogeneity/heterogeneity in sufficient detail: Boneau (1960); Levine and Dunlap (1982); MacDonald (1999); Myers and Well (1995); Rasmussen (1985); Rasmussen (1986); Schneider and Penfield (1997); Zimmerman (1987); and Zimmerman and Zumbo (1992). Together, these studies resulted in a total of 231 "cases", in which each case was based upon from 1,000 to 10,000 simulated data sets (i.e., replications). The studies are further described in Table 1. Also, we found but did not include the data from three studies (Hsu, 1938; Pratt, 1964; Scheffe, 1959) which reported an additional 18 cases. These cases differed from the Monte Carlo simulation data because they consisted of exact mathematical results for specific heterogeneity conditions. They did, however, provide information about the general pattern of results that we should expect from our regression analysis.

Insert Table 1 about here

As can be seen from the values describing the type and extent of normality violation presented in Table 1, several of the simulation studies probed what happens to alpha when assumption violations are quite extreme. In part, this likely reflects a desire to determine the upper and lower boundaries of the discrepancy between nominal and actual levels of alpha. Furthermore, the simulation studies that we drew upon were typically performed to demonstrate the superiority of some alternative to the t-test—the performance of the alternatives was generally highlighted best under conditions of extreme assumption violation. For example, the variance ratios used in the simulations ranged from 1:36 (Schneider & Penfiled, 1997) to 100:1 (Myers & Well, 1995), and the normality violations used in some studies were quite severe (e.g., kurtosis values that were below—1, skew values of 6 or above). On the other hand, the sample size ratios used in the studies were typically much more realistic in the sense of being likely to occur an actual data collection situation, although we note that the sample sizes and total N's might well be considered small for typical research studies in some topic areas. This issue potentially had some implications for our regression estimates, and as will be described in more detail shortly, resulted in a decision to restrict our analysis to those simulation cases in which the violations were within a range at least somewhat consistent with naturally-occurring conditions. We also return to this issue with a recommendation in our conclusions.

Accumulation and coding of results. Each of the previously-described studies was carefully inspected in order to identify and code at minimum the following information for each set of relevant simulations.

1. A variance ratio, which was defined as the variance of Population 1 divided by the variance of Population 2. This variable was designated VARRATIO. The homogeneity assumption is met when the variance ratio has a value of 1.
2. A sample size ratio, defined as the sample size selected from Population 1 divided by the sample size selected from Population 2. This variable was labeled SAMRATIO. Values were coded such that the Population 1 sample size was always equal to, or larger than, Population 2 sample size, and if necessary, the reported variance ratio described above was reflected to maintain a consistent ordering.
3. The total sample size (combined group sizes) was called TOTSAM.
4. Violations of the normality assumption were coded in multiple ways. First, two variables were created, one for each population, and information about the specific form of each population was recorded. The options noted were: normal, mixed

normal, exponential, rectangular (uniform), lognormal, chi-square, and various specific values of skew/kurtosis.

5. A set of four dummy-coded variables was created in which a value of "1" was assigned if the populations were: exponential (EXPO), rectangular (RECT), lognormal (LOGNORM), or mixed normal (MIXED), and a value of zero was assigned if not. In the few studies in which exact values of skew and kurtosis were reported, dummy codes were assigned so that case was treated as an example of the most representative form of distribution listed above. Cases drawn from a normal distribution were thus coded "0000" on this set of dummy variables.
6. Normality violation was also captured in a very simple manner with a dummy-coded variable which either had a value of "1" (both population distributions were normal) or "0"(at least one population distribution was non-normal). This variable was labeled BOTHNORMAL.
7. Distinctions between the nominal Type I error rate in the case were captured with three dummy-coded variables named ALPHA01, ALPHA05, and ALPHA10. Note that for purposes of the regression analysis, only two of these variables were used in a given model—the third would only provide redundant information.
8. The dependent variable in the regression model was the observed Type I error rate (across all samples in the relevant simulation). This variable was called ACTUAL, to indicate that it represented the actual alpha level across samples with a given type and degree of assumption violation.

In addition, a number of interaction terms were created, consisting of the following:

1. Two-way interaction effects:
    a. SAMVAR: an interaction between the ratio of the sample sizes (SAMRATIO) and the ratio of the population variances (VARRATIO).

b. SAMTOT: an interaction between the ratio of the sample sizes (SAMRATIO) and the total sample size (TOTSAM).

c. VARTOT: between the ratio of the population variances and sample sizes (SAMRATIO) and the total sample size (TOTSAM).

2. The three-way interaction effect among the ratio of the sample sizes (SAMRATIO), the ratio of the population variances (VARRATIO), and the total sample size (TOTSAM) was labeled (SAVATOT).

Thus the initial regression model that we tested was constructed as follows (McNeil, Newman, & Kelly, 1996):

$$\begin{aligned} \text{ACTUAL} = \; & a_0 + b1(\text{ALPHA01}) + b_2(\text{ALPHA10}) + \\ & b_3(\text{MIXED}) + b_4(\text{LOGNORM}) + b_5(\text{EXPO}) + b_6(\text{RECT}) + \\ & b_7(\text{VARRATIO}) + b_8(\text{TOTSAM}) + b_9(\text{SAMRATIO}) + \\ & b_{10}(\text{VARTOT}) + b_{11}(\text{SAMTOT}) + b_{12}(\text{SAMVAR}) + \\ & b_{13}(\text{SAVATOT}) + e \end{aligned}$$

where, a_0, and b_1 through b_{13} represent the estimated intercept and slope parameters, respectively.

Results

The model displayed above was first tested using the data on the 231 MC simulation cases. For this and other regression analyses, each analysis was performed twice, using both an ordinary least squares (OLS) and a weighted least squares (WLS) estimation procedure. In the WLS estimation, the number of replications per case was used as the weighting factor.

Our initial results showed that cases with very extreme variance ratios tended to emerge as outliers in the regression analysis, and they also seemed unrepresentative of the real life situations in which most research would be conducted. Thus we decided to further restrict the sample to those cases in which the variance ratio was between .05 and 20 (i.e., ratio in which the largest to smallest sample sizes had a variance ratios ranging from 1:20 to 20:1). This decision had the effect of eliminating all 54 cases from the Schneider and Penfield (1997) study, as well as a few cases from other studies. In addition, during analysis six extreme outlier values were identified. This data screening resulted in a useable sample of N=158 cases. Finally, the parameter estimates for the total sample size variable, along with all of its interactions were low magnitude and not statistically significant. Furthermore, their presence did not seem to affect other parameter estimates appreciably. Therefore these variables were trimmed from the final model.

The results produced by the OLS and WLS analyses of the trimmed multiple linear regression model are listed in Tables 2 and 3, respectively. The R^2 values of .877 for the OLS model, and .907 for the WLS model, were statistically significant at the .001 level. (The adjusted R^2 values for these two models were .870 and .902, respectively.) An inspection of the coefficients and their accompanying significance tests for both models leads to parallel conclusions for eight of the nine predictor variables. First, and not surprisingly, the nominal alpha level (which drives the choice of the criterion value for t used to determine statistical significance), as reflected in the two dummy codes ALPHA01 and ALPHA10, showed a significant relationship ($p < .001$) to the observed alpha under both estimation techniques. If this relationship had not held, we would have suspected a horribly misspecified model or some egregious coding errors! The parameter estimates indicate that when nominal alpha was .01, the value of observed alpha was lowered by approximately .03, and when it was .10, the observed alpha was increased by approximately .05.

Insert Tables 2 and 3 about here

Second, three of the four dummy variables coding for normality violations showed parallel results in the two estimations. Studies in which at least one of the samples was drawn from either a mixed normal or a log-normal distribution showed a significant deflation of alpha by about .01. And, both estimation methods suggested that the rectangular distributions studied did not show a significant effect on the observed alpha. However, the OLS estimation suggested that the exponential distributions did *not* have a significant effect on observed alpha, while the WLS results suggested that they did ($p=.004$), and the parameter estimates for the regression coefficient differed by an order of magnitude (.01 versus .001) for the two estimation methods.

Finally, the effects of sample size and variance heterogeneity were similar across the two estimation methods. They showed significant positive coefficients for the main effects of sample size ratio and variance ratio, and a significant negative coefficient for the interaction of the sample size and variance ratios (all $p < .001$). This result was consistent with previous research which suggests that the observed alpha is deflated when smaller sample sizes are associated with smaller variances.

As previously mentioned, the purpose of constructing and analyzing such a model was to produce estimated regression coefficients that practitioners could use to estimate the actual alpha level for an independent groups t-test conducted in their study. Thus, we also considered whether our models were likely to yield defensible estimates of observed alpha. For example, one would expect that the prediction for a situation in which the nominal alpha was .05, the populations distributions are both normal, sample sizes are equal, and variances are homogeneous, that the regression model would return an estimated observed alpha of .05. Both of our models perform relatively well on this task, with a predicted value of .049 for the OLS estimation and a predicted values of .047 for the WLS estimation.

To illustrate a more complex situation, assume that the following sample and population characteristics existed in a study in which an independent groups t-test was conducted: (a) the sample size ratio was 2.0, (b) the variance ratio was .5, (c) the distributions of both populations were assumed to be mixed normal, and (d) the nominal alpha level was .05. Under these conditions, the estimated regression coefficients and values of the predictor variables are as follows:

1. OLS coefficient values:

   prediction of observed alpha = .0413 -.0136(MIXED=1) +
   .0073(SAMRATIO=2) + .0052(VARRATIO=.5) -
   .0048(SAMVAR=1) = <u>.0401</u>

2. WLS coefficient values:

   prediction of observed alpha = .0371 -.0154(MIXED=1) +
   .0096(SAMRATIO=2) + .0053(VARRATIO=.5) -
   .0046(SAMVAR=1) = <u>.0389</u>

These two estimated actual alpha values suggest that for this particular situation, the actual alpha value is not likely to exceed the nominal alpha value, nor is it so deflated that power might become a serious concern. Even when there is more pronounced heterogeneity of variance combined with unequal sample sizes, the observed value of alpha may be influenced less than one would suspect, given warnings to avoid the t-test under these situations. For example, consider a case where both population distributions are normal, but the sample sizes are quite discrepant with a ratio of 10:1 and the variances are also quite discrepant with a ratio of 1:10. The OLS and WLS models in this situation would predict an observed alpha of .11 (OLS) or .13 (WLS). Although alpha is clearly inflated in this situation, the researcher may yet find rejecting the null hypothesis poses an acceptable level of risk.

Conclusions and Implications

Much good statistical research has gone into demonstrating how violations of assumptions can produce misleading results, and into the development of specialized procedures that can be applied to correct estimates and probability levels when violations occur. However, if the focus of the analysis is correct decision-making with respect to retaining or rejecting the null hypothesis, then one might ask if these specialized procedures are always necessary? The answer to this question depends in part upon several factors. First, how serious do the violations tend to be? For example, if normality violations are common, are we typically considering the effects of a slight violation of normality, or drastically non-normal distributions? Second, are the effects of these violations likely to inflate or deflate probability values? Finally, although an ideal preference would be to use the appropriate statistical procedure to correct for a given violation of an assumption, how much of an adjustment is typically required to maintain the nominal alpha? It is a procedure to address this third question that the current paper investigates.

Our regression results suggest that the independent groups t-test is fairly robust. If this doesn't find this generality acceptable, then a more precise estimate of the risk of an inflated alpha can be made, if violations of assumptions can be determined. Results could then be presented and interpreted in light of this estimate of the actual alpha level.

However, a few caveats are in order. First, this analysis has considered the effects of violating assumptions on the Type I error rate only. It may be equally or more important in a given context to determine the extent to which statistical power (e.g., the chances of avoiding a Type II error) is compromised. We suggest that the same general technique that we have presented here could be used to address this question also, by simply cumulating the results from Monte Carlo studies which have determined the effects of assumption violations on power. An important consideration in the development of a predictive model for effects on power would be some information on the effect size of a true difference between the group means.

Another issue that bears further investigation is the issue of gaining sufficient information about the values of the predictors in the regression models. For example, our coding for the normality violations was fairly crude. A better scheme would employ more precise information about the population values of skew and kurtosis. Unfortunately, this detailed information was not available for all of the studies in our sample. This could result in more precise parameter estimates for the regression models.

The development of such predictive models also suggests a possible function for future simulation studies, which might be conducted specifically for the purpose of systematically generating cases for a regression analysis. We note, for example, that when the purpose of the Monte Carlo simulations was to generate examples of the superiority of alternatives to the t-test, researchers quite rightly looked at unusual but interesting situations involving extreme combinations of non-normality, unbalanced sample sizes and heterogeneity of variance. However, we wonder if for our purposes we would be well-served by simulations generating many intermediate values of cases that were mildly non-normal, unbalanced, and/or heterogeneous.

Finally, we presented the results of both OLS and WLS estimation procedures, motivated in part by the work of Harwell (1992, Harwell et al., 1992). The weighting employed in the WLS analysis made sense to the extent that it more heavily weighted simulation results that depended upon a greater number of replications—other things being equal-- one might expect that such results would be more accurate. However, in another sense the weighted analysis could be misleading—one would also want the weights to reflect the extent to which the attendant violation assumption was frequent or infrequent in the relevant population of studies. Again, the creation of a set of Monte

Carlo simulations which takes this latter consideration into account might lead to a more representative data set, which in turn generates more accurate regression coefficients.

In sum, we have taken an initial step towards developing a statistical tool that researchers may find useful in practice, when they must balance the costs and benefits of relying on a statistical test that may be more familiar and understood, and easier to perform, but also known to result in inflated or deflated Type I errors when one cannot meet all of the underlying statistical assumptions. We suggest that in many situations, the t-test will prove adequately robust to error violations.

References

Boneau, C. A. (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin, 57*, 49-64.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics, 25*, 290-302.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics, 25*, 484-498.

Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research, 52*, 291-302.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research, 42*, 237-288.

Harwell, M. R. (1992). Summarizing Monte Carlo results in methodological research. *Journal of Educational Statistics, 17*, 297-313.

Harwell, M. R., Rubenstein, E. M., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. *Journal of Educational Statistics, 17*, 315-339.

Hays, W. L. (1994). *Statistics* (5^{th} Edition). Forth Worth, IN: Harcourt Brace College Publishers.

Hsu, P. L. (1938). On the best unbiased quadratic estimate of the variance. *Statistical Research Memoirs, 2*, 91-104.

Huberty, C. J. (1987). On statistical testing. *Educational Researcher, 16*, 4-9.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in the analysis of variance. *Psychological Bulletin, 99,* 422-431.

Levine, D. W., & Dunlap, W. P. (1982). Power of the F test with skewed data: Should one transform or not? *Psychological Bulletin, 92,* 272-280.

Maxwell, S. E., & Delaney, H. D. (2000). *Designing experiments and analyzing data: A model comparison approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

MacDonald, P. (1999). Power, Type I, and Type II error rates of parametric and nonparametric statistical tests. *The Journal of Experimental Education, 67,* 367-379.

McNeil, K., Newman, I., & Kelly, F. J. (1996). *Testing research hypotheses with the general linear model.* Carbondale, IL: Southern Illinois University Press.

Myers, J. L., & Well, A. D. (1995). *Research design and statistical analysis.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Norton, D. W. (1952). An empirical investigation of the effects of nonnormality and heterogeneity upon the F-test of analysis of variance. Unpublished doctoral dissertation, University of Iowa, Iowa City.

Pearson, E. S. (1931). The analysis of variance in the case of non-normal variation. *Biometrika, 23,* 114-133.

Pratt (1964) as reported in Scheffe (1959).

Rasmussen, J. L. (1985). The power of student's t and Wilcoxon statistics. *Evaluation Review, 9,* 505-510.

Rasmussen, J. L. (1986). An evaluation of parametric and non-parametric tests on modified and non-modified data. *Journal of Mathematical and Statistical Psychology, 39,* 213-220.

Scheffe, H. (1959). *The analysis of variance.* New York: John Wiley.

Schneider, P. J., & Penfield, D. A. (1997). Alexander and Govern's approximation: Providing an alternative to ANOVA under variance heterogeneity. *The Journal of Experimental Education, 65*, 271-286.

Tiku, M. L. (1964). Approximating the general non-normal variance-ratio sampling distributions. *Biometrika, 51*, 83-95.

Young, R. K., & Veldman, D. J. (1963). Heterogeneity and skewness in analysis of variance. *Perceptual and Motor Skills, 16*, 588.

Zimmerman, D. W. (1987). Comparative power of student t test and Mann-Whitney U test for unequal sample sizes and variances. *Journal of Experimental Education, 55*, 171-174.

Zimmerman, D. W., & Zumbo, B. D. (1992). Parametric alternatives to the student t test under violation of normality and homogeneity of variance. *Perceptual and Motor Skills, 74*, 835-844.

Table 1

Case Source and Descriptive Information

Source	Number of Cases	Nominal alpha	N	Ratio n_1/n_2	Variance Ratio	Distribution	Reps/Case
Simulation Studies:							
Boneau (1960), Table 1	40	.01 .05	10 20 30 50	1 3	.25 1 4	Normal Exponential (+ skew) Rectangular (kurtotic)	1,000
Levine & Dunlap (1982), Table 1	24	.01 .05 .10	8 16 32 64	1	1	Normal Lognormal	10,000
MacDonald (1999), Table 1	24	.10	10 20 40 60	1 2 3	1	Normal Exponential Mixed Normal: 80/90% N(0,1) +20/10% N(0,20)	10,000
Myers & Well (1995), Table 3.2	28	.01 .05	10 15 25 30 50	1 1.5 2	.25 4 16 100	Normal	2,000
Rasmussen (1985), Tables 1 & 2	12	.01 .05	12 36	1 3	1	Mixed Normal: 95% N(0,1) + 5% N(0,10)	5,000

Rasmussen (1986), Tables 1-4	16	.01 .05	12 36	1 3	1	Normal Mixed Normal: 87.5% N(0,1) + 12.5% N(33,10) (+ skew, bimodal)	1,000
Schneider & Penfield (1997), Table 2	54	.05	20 25	1 4	.03 .06 16 36	Normal Skewed .75, 1.5, 1.75 Kurtotic -1, 1.5, 3 3.75 Chi-Square	10,000
Zimmerman (1987), Table 1	18	.05	20	1 4	.04 1 25	Normal	2000
Zimmerman & Zumbo (1992), Figure 1 and Table 1	15	.05	32	1 3	.11 .16 .25 .44 1 2.25 3 4 6.25 9	Normal Mixed Normal	10,000

Total		.01	8	1	.03	Normal	
Including Schneider & Penfield (1997)	231	.05	10	1.5	.04	Mixed Normal	
Dropping Schneider & Penfield (1997)	177	.10	12	1.67	.06	Lognormal	
			14	2	.11	Exponential	
			15	3	.16	Rectangular	
			16	4	.25	Add□l Skew/Kurtot	
			20	5	.44		
			25		1		
			30		2.25		
			32		3		
			36		4		
			40		6.25		
			50		9		
			60		16		
			64		25		
					36		
					100		

Table 2

Multiple Linear Regression Model (OLS) Results for Simulation Cases, Outliers Removed (N=158), All Levels of Nominal Alpha

Variables	Coefficient	SE	t	p
Constant	.0413	.003	13.83	<.001
ALPHA01	-.0344	.002	-16.30	<.001
ALPHA10	.0485	.003	18.53	<.001
MIXED	-.0136	.003	-5.42	<.001
LOGNORM	-.0119	.004	-3.16	.002
EXPO	-.0015	.003	-0.50	.616
RECT	.0055	.003	1.60	.112
SAMRATIO	.0073	.001	5.19	<.001
VARRATIO	.0052	.001	6.60	<.001
SAMVAR	-.0048	.001	-8.42	<.001

Note. Model 1: $R^2 = .877$, adj. $R^2 = .870$, $F(9,148) = 117.384$, $p<.001$.

Table 3

Multiple Linear Regression Model (WLS) Results for Simulation Cases, Outliers Removed (N=158), All Levels of Nominal Alpha

Variables	Coefficient	SE	t	p
Constant	.0371	.003	12.30	<.001
ALPHA01	-.0335	.003	-13.20	<.001
ALPHA10	.0507	.002	22.20	<.001
MIXED	-.0154	.002	-6.47	<.001
LOGNORM	-.0114	.003	-4.05	<.001
EXPO	-.0100	.003	-2.96	.004
RECT	.0089	.007	1.28	.202
SAMRATIO	.0096	.001	6.75	<.001
VARRATIO	.0053	.001	6.31	<.001
SAMVAR	-.0046	<.001	-9.53	<.001

Note. Model 1: $R^2 = .907$, adj. $R^2 = .902$, $F(9,148) = 101.784$, $p<.001$.

REPRODUCTION RELEASE

(Specific Document)

TM034970

I. DOCUMENT IDENTIFICATION:

Title:	Development of a Regression Model for Estimating the Effects of Assumption Violations on Type I Error Rates in the Student's t Test: Implications for Practitioners
Author(s):	Isadore Newman, Rosalie J. Hall, and John W. Fraas
Corporate Source: The University of Akron	Publication Date: April 21, 2003

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please →

Signature: John W. Fraas	Printed Name/Position/Title: John W. Fraas, Ph.D. Trustees' Professor	
Organization/Address: Ashland University, 401 College Avenue Ashland, Ohio, 44805	Telephone: 419-289-5930	FAX: 419-289-5910
	E-Mail Address: jfraas@ashland.edu	Date: 5/20/03

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse: University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com